

**Analisa Sentimen *Tweet* Berbahasa Indonesia dengan
Menggunakan Metode Pembobotan Hybrid TF-IDF pada
Topik Transportasi Online**

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
Teknik Informatika Universitas Muhammadiyah Malang



Oleh :

Sari Wahyunita
201310370311014

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MALANG
2017**

LEMBAR PERSETUJUAN

Analisa Sentimen *Tweet* Berbahasa Indonesia dengan Menggunakan Metode Pembobotan Hybrid TF-IDF pada Topik Transportasi Online

Diajukan Untuk Memenuhi
Persyaratan Guna Meraih Gelar Sarjana Strata 1
Teknik Informatika Universitas Muhammadiyah Malang

Sari Wahyunita

201310370311014

Menyetujui

Dosen 1



Yufis Azhar, S.Kom., M.Kom

NIP. 108.1410.0544

Dosen 2



Nur Hayatin, S.ST

NIDN. 108.0907.0476

LEMBAR PENGESAHAN

Analisa Sentimen *Tweet* Berbahasa Indonesia dengan Menggunakan Metode Pembobotan *Hybrid TF-IDF* pada Topik Transportasi Online

Diajukan Untuk Memenuhi
Persyaratan Guna Meraih Gelar Sarjana Strata 1
Teknik Informatika Universitas Muhammadiyah Malang

Sari Wahyunita
201310370311014

Menyetujui

Penguji 1

Galih Wasis W, S.Kom., M.Cs

NIP. 108.1419.0540

Penguji 2

Christian S. K. Aditya, M.Kom

NIP. 180.3270.21991

Mengetahui,

Ketua Jurusan Teknik Informatika

Gita Indah M, ST., M. Kom.

NIP. 108.0611.0442

DAFTAR ISI

ABSTRAK	i
ABSTRACT	ii
LEMBAR PERSETUJUAN.....	iii
LEMBAR PENGESAHAN	iv
LEMBAR PERNYATAAN	v
LEMBAR PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian	3
1.4. Batasan Masalah.....	3
1.5. Metode Penelitian.....	4
1.6. Sistematika Penulisan	5
BAB II LANDASAN TEORI.....	6
2.1. Twitter dan Twitter API.....	6
2.2. Analisis Sentimen	7
2.3. Preprocessing	8
2.4. Library Sastrawi.....	9
2.5. Pembobotan Hybrid TF-IDF (Term Frequency-Inverse Document Frequency).....	10
2.6. Klasifikasi kNN (k-Nearest Neighbor)	11
2.7. Cosine Similarity.....	12
2.8. Pengujian.....	13
BAB III ANALISA DAN PERANCANGAN SISTEM.....	14
3.1. Analisa Data	14
3.2. Kebutuhan Software dan Hardware yang Digunakan.....	14
3.3. Pengumpulan Data atau Crawling.....	14

3.4. Analisa dan Perancangan	16
3.5. Pengujian.....	25
BAB IV IMPLEMENTASI DAN PENGUJIAN	29
4.1. Crawling Data	29
4.2. Preprocessing	32
4.3. Pembobotan Hybrid TF-IDF	34
4.4. Klasifikasi kNN.....	35
4.5. Pengujian.....	36
BAB V KESIMPULAN DAN SARAN.....	43
5.1. Kesimpulan	43
5.2. Saran.....	43
DAFTAR PUSTAKA	44
LAMPIRAN.....	47



DAFTAR PUSTAKA

- [1] M. Idris, "Pendapatan Operator Taksi Menurun Hingga 20% dengan Kehadiran Uber dan GrabCar," *detiknews*, 15 Maret 2016. [Online]. Available: <http://news.detik.com/berita/3165540/pendapatan-operator-taksi-menurun-hingga-20-dengan-kehadiran-uber-dan-grabcar>. [Accessed 9 Mei 2017].
- [2] Twitter, "Pencapaian Twitter," 30 Juni 2016. [Online]. Available: <https://about.twitter.com/id/company/press/milestones>. [Accessed 11 Maret 2017].
- [3] Statistika, "Statistika: Number of Active Twitter Users in Leading Markets as of May 2016 (in Millions)," Mei 2016. [Online]. Available: <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>. [Accessed 12 Maret 2017].
- [4] J. Weng, E.-P. Lim and J. Jiang, "TwitterRank: Finding Topic-sensitive Influential Twitters," *WSDM'10*, pp. 261-270, 2010.
- [5] R. Wati, "Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes," *Evolusi*, vol. 4, pp. 26-32, 2016.
- [6] D. H. Wahid and A. SN, "Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity," *IJCCS*, vol. 10, pp. 207-218, 2016.
- [7] J. Riany, M. Fajar and M. P. Lukman, "Penerapan Deep Sentiment Analysis pada Angket Penilaian Terbuka Menggunakan K-Nearest Neighbor," *jurnal sisfo*, vol. 06, pp. 147-156, 2016.
- [8] G.Vinodhini and RM.Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 6, pp. 282-292, 2012.
- [9] B. Liu, *Sentiment Analysis and Opinion Mining*, San Rafael: Morgan&Claypool Publishers, 2012.

- [10] C. R. Fink, D. S. Chou, J. J. Kopecky and A. J. Llorens, "Coarse- and Fine-Grained Sentiment Analysis of Social Media Text," *Johns Hopkins APL Technical Digest*, vol. 30, no. Number 1, pp. 22-30, 2011.
- [11] N. D. Putranti and E. Winarko, "Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine," *IJCCS*, vol. 8, pp. 91-100, 2014.
- [12] M. Adriani, J. Asian, B. Nazief, S. Tahaghoghi and H. E. Williams, "Stemming Indonesian: A Confix Stripping Approach," *ACM Transactions on Asian Language Information Processing*, vol. 6, pp. 13-33, 2007.
- [13] J. Asian, *Effective Techniques for Indonesian Text Retrieval*, Melbourne, 2007.
- [14] A. Z. Arifin, I. P. A. K. Mahendra and H. T. Ciptaningtyas, "ENHANCED CONFIX STRIPPING STEMMER AND ANTS ALGORITHM FOR CLASSIFYING NEWS DOCUMENT IN INDONESIAN LANGUAGE," *The International Conference on Information & Communication Technology and Systems*, vol. 5, pp. 149-158, 2009.
- [15] A. D. Tahitoe and D. Purwitasari, "Implementasi Modifikasi Enhanced Confix Stripping Stemmer untuk Bahasa Indonesia dengan Metode Corpus Based Stemming," *Institut Teknologi Sepuluh November (ITS) - Surabaya*, pp. 1-15, 2010.
- [16] B. Sharifi, M.-A. Hutton and J. K. Kalita, "Experiments in Microblog Summarization," *IEEE Second International Conference*, pp. 49-56, 2010.
- [17] B. P. Sharifi, D. I. Inouye and J. K. Kalita, "Summarization of Twitter Microblogs," *The Computer Journal*, vol. 57, pp. 378-402, 2013.
- [18] M. I. Sikki, "PENGENALAN WAJAH MENGGUNAKAN K-NEAREST NEIGHBOUR DENGAN PRAPROSES TRANSFORMASI WAVELET," *Jurnal Paradigma*, vol. X, pp. 159-172, 2009.
- [19] O. Nurdiana, Jumadi and D. Nursantika, "PERBANDINGAN METODE COSINE SIMILARITY DENGAN METODE JACCARD SIMILARITY PADA APLIKASI PENCARIAN TERJEMAH AL-QUR'AN DALAM BAHASA INDONESIA," *JOIN*, vol. I, pp. 59-63, 2016.

- [20] H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *J Pikel*, vol. 1, pp. 65-76, 2013.

